Causation

David Lewis

The Journal of Philosophy, Volume 70, Issue 17, Seventieth Annual Meeting of the American Philosophical Association Eastern Division (Oct. 11, 1973), 556-567.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The Journal of Philosophy is published by The Journal of Philosophy, Inc.. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/jphil.html.

The Journal of Philosophy
©1973 The Journal of Philosophy, Inc.

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact jstor-info@umich.edu.

©2001 JSTOR

http://www.jstor.org/
Mon Nov 12 19:16:15 2001

CAUSATION *

HUME defined causation twice over. He wrote "we may define a cause to be *an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second.* Or, in other words, *where, if the first object had not been, the second never had existed.*"[1]

Descendants of Hume's first definition still dominate the philosophy of causation: a causal succession is supposed to be a succession that instantiates a regularity. To be sure, there have been improvements. Nowadays we try to distinguish the regularities that count—the "causal laws"—from mere accidental regularities of succession. We subsume causes and effects under regularities by means of descriptions they satisfy, not by over-all similarity. And we allow a cause to be only one indispensable part, not the whole, of the total situation that is followed by the effect in accordance with a law. In present-day regularity analyses, a cause is defined (roughly) as any member of any minimal set of actual conditions that are jointly sufficient, given the laws, for the existence of the effect.

More precisely, let C be the proposition that c exists (or occurs) and let E be the proposition that e exists. Then c causes e, according to a typical regularity analysis,[2] iff (1) C and E are true; and (2) for some nonempty set $\mathfrak{L}$ of true law-propositions and some set $\mathfrak{F}$ of true propositions of particular fact, $\mathfrak{L}$ and $\mathfrak{F}$ jointly imply $C \supset E$, although $\mathfrak{L}$ and $\mathfrak{F}$ jointly do not imply E and $\mathfrak{F}$ alone does not imply $C \supset E$.[3]

Much needs doing, and much has been done, to turn definitions like this one into defensible analyses. Many problems have been overcome. Others remain: in particular, regularity analyses tend to confuse causation itself with various other causal relations. If c belongs to a minimal set of conditions jointly sufficient for e, given the laws,

* To be presented in an APA symposium on Causation, December 28, 1973; commentators will be Bernard Berofsky and Jaegwon Kim; see this JOURNAL, this issue, pp. 568–569 and 570–572, respectively.

I thank the American Council of Learned Societies, Princeton University, and the National Science Foundation for research support.

[1] *An Enquiry concerning Human Understanding*, Section VII.

[2] Not one that has been proposed by any actual author in just this form, so far as I know.

[3] I identify a *proposition*, as is becoming usual, with the set of possible worlds where it is true. It is not a linguistic entity. Truth-functional operations on propositions are the appropriate Boolean operations on sets of worlds; logical relations among propositions are relations of inclusion, overlap, etc. among sets. A sentence of a language *expresses* a proposition iff the sentence and the proposition are true at exactly the same worlds. No ordinary language will provide sentences to express all propositions; there will not be enough sentences to go around.

then c may well be a genuine cause of e. But c might rather be an effect of e: one which could not, given the laws and some of the actual circumstances, have occurred otherwise than by being caused by e. Or c might be an epiphenomenon of the causal history of e: a more or less inefficacious effect of some genuine cause of e. Or c might be a preempted potential cause of e: something that did not cause e, but that would have done so in the absence of whatever really did cause e.

It remains to be seen whether any regularity analysis can succeed in distinguishing genuine causes from effects, epiphenomena, and preempted potential causes—and whether it can succeed without falling victim to worse problems, without piling on the epicycles, and without departing from the fundamental idea that causation is instantiation of regularities. I have no proof that regularity analyses are beyond repair, nor any space to review the repairs that have been tried. Suffice it to say that the prospects look dark. I think it is time to give up and try something else.

A promising alternative is not far to seek. Hume's "other words" —that if the cause had not been, the effect never had existed—are no mere restatement of his first definition. They propose something altogether different: a counterfactual analysis of causation.

The proposal has not been well received. True, we do know that causation has something or other to do with counterfactuals. We think of a cause as something that makes a difference, and the difference it makes must be a difference from what would have happened without it. Had it been absent, its effects—some of them, at least, and usually all—would have been absent as well. Yet it is one thing to mention these platitudes now and again, and another thing to rest an analysis on them. That has not seemed worth while.[4] We have learned all too well that counterfactuals are ill understood, wherefore it did not seem that much understanding could be gained by using them to analyze causation or anything else. Pending a better understanding of counterfactuals, moreover, we had no way to fight seeming counterexamples to a counterfactual analysis.

But counterfactuals need not remain ill understood, I claim, unless we cling to false preconceptions about what it would be like to understand them. Must an adequate understanding make no reference to unactualized possibilities? Must it assign sharply determinate truth conditions? Must it connect counterfactuals rigidly to covering laws? Then none will be forthcoming. So much the worse for those standards of adequacy. Why not take counterfactuals at face value:

[4] One exception: Aardon Lyon, "Causality," *British Journal for Philosophy of Science*, XVIII, 1 (May 1967): 1–20.

as statements about possible alternatives to the actual situation, somewhat vaguely specified, in which the actual laws may or may not remain intact? There are now several such treatments of counterfactuals, differing only in details.[6] If they are right, then sound foundations have been laid for analyses that use counterfactuals.

In this paper, I shall state a counterfactual analysis, not very different from Hume's second definition, of some sorts of causation. Then I shall try to show how this analysis works to distinguish genuine causes from effects, epiphenomena, and preempted potential causes.

My discussion will be incomplete in at least four ways. Explicit preliminary settings-aside may prevent confusion.

1. I shall confine myself to causation among *events*, in the everyday sense of the word: flashes, battles, conversations, impacts, strolls, deaths, touchdowns, falls, kisses, and the like. Not that events are the only things that can cause or be caused; but I have no full list of the others, and no good umbrella-term to cover them all.

2. My analysis is meant to apply to causation in particular cases. It is not an analysis of causal generalizations. Presumably those are quantified statements involving causation among particular events (or non-events), but it turns out not to be easy to match up the causal generalizations of natural language with the available quantified forms. A sentence of the form "C-events cause E-events," for instance, can mean any of

(a) For some c in C and some e in E, c causes e.

(b) For every e in E, there is some c in C such that c causes e.

(c) For every c in C, there is some e in E such that c causes e.

not to mention further ambiguities. Worse still, 'Only C-events cause E-events' ought to mean

(d) For every c, if there is some e in E such that c causes e, then c is in C.

if 'only' has its usual meaning. But no; it unambiguously means (b) instead! These problems are not about causation, but about our idioms of quantification.

3. We sometimes single out one among all the causes of some event and call it "the" cause, as if there were no others. Or we single out a few as the "causes," calling the rest mere "causal factors" or "causal conditions." Or we speak of the "decisive" or "real" or "principal" cause. We may select the abnormal or extraordinary

[6] See, for instance, Robert Stalnaker, "A Theory of Conditionals," in Nicholas Rescher, ed., *Studies in Logical Theory* (Oxford: Blackwell, 1968); and my *Counterfactuals* (Oxford: Blackwell, 1973).

causes, or those under human control, or those we deem good or bad, or just those we want to talk about. I have nothing to say about these principles of invidious discrimination.[6] I am concerned with the prior question of what it is to be one of the causes (unselectively speaking). My analysis is meant to capture a broad and nondiscriminatory concept of causation.

4. I shall be content, for now, if I can give an analysis of causation that works properly under determinism. By determinism I do not mean any thesis of universal causation, or universal predictability-in-principle, but rather this: the prevailing laws of nature are such that there do not exist any two possible worlds which are exactly alike up to some time, which differ thereafter, and in which those laws are never violated. Perhaps by ignoring indeterminism I squander the most striking advantage of a counterfactual analysis over a regularity analysis: that it allows undetermined events to be caused.[7] I fear, however, that my present analysis cannot yet cope with all varieties of causation under indeterminism. The needed repair would take us too far into disputed questions about the foundations of probability.

COMPARATIVE SIMILARITY

To begin, I take as primitive a relation of *comparative over-all* similarity among possible worlds. We may say that one world is *closer to actuality* than another if the first resembles our actual world more than the second does, taking account of all the respects of similarity and difference and balancing them off one against another.

(More generally, an arbitrary world w can play the role of our actual world. In speaking of our actual world without knowing just which world is ours, I am in effect generalizing over all worlds. We really need a three-place relation: world w_1 is closer to world w than world w_2 is. I shall henceforth leave this generality tacit.)

I have not said just how to balance the respects of comparison against each other, so I have not said just what our relation of comparative similarity is to be. Not for nothing did I call it primitive. But I have said what *sort* of relation it is, and we are familiar with relations of that sort. We do make judgments of comparative over-all similarity—of people, for instance—by balancing off many re-

[6] Except that Morton G. White's discussion of causal selection, in *Foundations of Historical Knowledge* (New York: Harper & Row, 1965), pp. 105–181, would meet my needs, despite the fact that it is based on a regularity analysis.

[7] That this ought to be allowed is argued in G. E. M. Anscombe, *Causality and Determination: An Inaugural Lecture* (Cambridge: University Press, 1971); and in Fred Dretske and Aaron Snyder, "Causal Irregularity," *Philosophy of Science*, XXXIX, 1 (March 1972): 69–71.

spects of similarity and difference. Often our mutual expectations about the weighting factors are definite and accurate enough to permit communication. I shall have more to say later about the way the balance must go in particular cases to make my analysis work. But the vagueness of over-all similarity will not be entirely resolved. Nor should it be. The vagueness of similarity does infect causation, and no correct analysis can deny it.

The respects of similarity and difference that enter into the over-all similarity of worlds are many and varied. In particular, similarities in matters of particular fact trade off against similarities of law. The prevailing laws of nature are important to the character of a world; so similarities of law are weighty. Weighty, but not sacred. We should not take it for granted that a world that conforms perfectly to our actual laws is *ipso facto* closer to actuality than any world where those laws are violated in any way at all. It depends on the nature and extent of the violation, on the place of the violated laws in the total system of laws of nature, and on the countervailing similarities and differences in other respects. Likewise, similarities or differences of particular fact may be more or less weighty, depending on their nature and extent. Comprehensive and exact similarities of particular fact throughout large spatiotemporal regions seem to have special weight. It may be worth a small miracle to prolong or expand a region of perfect match.

Our relation of comparative similarity should meet two formal constraints. (1) It should be a weak ordering of the worlds: an ordering in which ties are permitted, but any two worlds are comparable. (2) Our actual world should be closest to actuality, resembling itself more than any other world resembles it. We do *not* impose the further constraint that for any set A of worlds there is a unique closest A-world, or even a set of A-worlds tied for closest. Why not an infinite sequence of closer and closer A-worlds, but no closest?

COUNTERFACTUALS AND COUNTERFACTUAL DEPENDENCE

Given any two propositions A and C, we have their *counterfactual* $A \mathrel{\square\!\!\rightarrow} C$: the proposition that if A were true, then C would also be true. The operation $\mathrel{\square\!\!\rightarrow}$ is defined by a rule of truth, as follows. $A \mathrel{\square\!\!\rightarrow} C$ is true (at a world w) iff either (1) there are no possible A-worlds (in which case $A \mathrel{\square\!\!\rightarrow} C$ is *vacuous*), or (2) some A-world where C holds is closer (to w) than is any A-world where C does not hold. In other words, a counterfactual is nonvacuously true iff it takes less of a departure from actuality to make the consequent true along with the antecedent than it does to make the antecedent true without the consequent.

We did not assume that there must always be one or more closest A-worlds. But if there are, we can simplify: $A \mathrel{\Box\!\!\rightarrow} C$ is nonvacuously true iff C holds at all the closest A-worlds.

We have not presupposed that A is false. If A is true, then our actual world is the closest A-world, so $A \mathrel{\Box\!\!\rightarrow} C$ is true iff C is. Hence $A \mathrel{\Box\!\!\rightarrow} C$ implies the material conditional $A \supset C$; and A and C jointly imply $A \mathrel{\Box\!\!\rightarrow} C$.

Let $A_1, A_2, \ldots$ be a family of possible propositions, no two of which are compossible; let $C_1, C_2, \ldots$ be another such family (of equal size). Then if all the counterfactuals $A_1 \mathrel{\Box\!\!\rightarrow} C_1$, $A_2 \mathrel{\Box\!\!\rightarrow} C_2$, $\ldots$ between corresponding propositions in the two families are true, we shall say that the C's *depend counterfactually* on the A's. We can say it like this in ordinary language: whether C_1 or C_2 or $\ldots$ depends (counterfactually) on whether A_1 or A_2 or $\ldots$.

Counterfactual dependence between large families of alternatives is characteristic of processes of measurement, perception, or control. Let $R_1, R_2, \ldots$ be propositions specifying the alternative readings of a certain barometer at a certain time. Let $P_1, P_2, \ldots$ specify the corresponding pressures of the surrounding air. Then, if the barometer is working properly to measure the pressure, the R's must depend counterfactually on the P's. As we say it: the reading depends on the pressure. Likewise, if I am seeing at a certain time, then my visual impressions must depend counterfactually, over a wide range of alternative possibilities, on the scene before my eyes. And if I am in control over what happens in some respect, then there must be a double counterfactual dependence, again over some fairly wide range of alternatives. The outcome depends on what I do, and that in turn depends on which outcome I want.[8]

CAUSAL DEPENDENCE AMONG EVENTS

If a family $C_1, C_2, \ldots$ depends counterfactually on a family $A_1, A_2, \ldots$ in the sense just explained, we will ordinarily be willing to speak also of causal dependence. We say, for instance, that the barometer reading depends causally on the pressure, that my visual impressions depend causally on the scene before my eyes, or that the outcome of something under my control depends causally on what I do. But there are exceptions. Let $G_1, G_2, \ldots$ be alternative possible laws of gravitation, differing in the value of some numerical constant. Let $M_1, M_2, \ldots$ be suitable alternative laws of planetary motion. Then the M's may depend counterfactually on the G's, but

[8] Analyses in terms of counterfactual dependence are found in two papers of Alvin I. Goldman: "Toward a Theory of Social Power," *Philosophical Studies*, XXIII (1972): 221–268; and "Discrimination and Perceptual Knowledge," presented at the 1972 Chapel Hill Colloquium.

we would not call this dependence causal. Such exceptions as this, however, do not involve any sort of dependence among distinct particular events. The hope remains that causal dependence among events, at least, may be analyzed simply as counterfactual dependence.

We have spoken thus far of counterfactual dependence among propositions, not among events. Whatever particular events may be, presumably they are not propositions. But that is no problem, since they can at least be paired with propositions. To any possible event e, there corresponds the proposition $O(e)$ that holds at all and only those worlds where e occurs. This $O(e)$ is the proposition that e occurs.[9] (If no two events occur at exactly the same worlds—if, that is, there are no absolutely necessary connections between distinct events—we may add that this correspondence of events and propositions is one to one.) Counterfactual dependence among events is simply counterfactual dependence among the corresponding propositions.

Let $c_1, c_2, \ldots$ and $e_1, e_2, \ldots$ be distinct possible events such that no two of the c's and no two of the e's are compossible. Then I say that the family $e_1, e_2, \ldots$ of events *depends causally* on the family $c_1, c_2, \ldots$ iff the family $O(e_1), O(e_2), \ldots$ of propositions depends counterfactually on the family $O(c_1), O(c_2), \ldots$. As we say it: whether e_1 or e_2 or . . . occurs depends on whether c_1 or c_2 or . . . occurs.

We can also define a relation of dependence among single events rather than families. Let c and e be two distinct possible particular

[9] Beware: if we refer to a particular event e by means of some description that e satisfies, then we must take care not to confuse $O(e)$, the proposition that e itself occurs, with the different proposition that some event or other occurs which satisfies the description. It is a contingent matter, in general, what events satisfy what descriptions. Let e be the death of Socrates—the death he actually died, to be distinguished from all the different deaths he might have died instead. Suppose that Socrates had fled, only to be eaten by a lion. Then e would not have occurred, and $O(e)$ would have been false; but a different event would have satisfied the description 'the death of Socrates' that I used to refer to e. Or suppose that Socrates had lived and died just as he actually did, and afterwards was resurrected and killed again and resurrected again, and finally became immortal. Then no event would have satisfied the description. (Even if the temporary deaths are real deaths, neither of the two can be *the* death.) But e would have occurred, and $O(e)$ would have been true. Call a description of an event e *rigid* iff (1) nothing but e could possibly satisfy it, and (2) e could not possibly occur without satisfy it. I have claimed that even such common-place descriptions as 'the death of Socrates' are nonrigid, and in fact I think that rigid descriptions of events are hard to find. That would be a problem for anyone who needed to associate with every possible event e a sentence $\phi(e)$ true at all and only those worlds where e occurs. But we need no such sentences—only propositions, which may or may not have expressions in our language.

events. Then *e depends causally* on *c* iff the family $O(e)$, $\sim O(e)$ depends counterfactually on the family $O(c)$, $\sim O(c)$. As we say it: whether *e* occurs or not depends on whether *c* occurs or not. The dependence consists in the truth of two counterfactuals: $O(c) \mathrel{\Box\!\!\rightarrow} O(e)$ and $\sim O(c) \mathrel{\Box\!\!\rightarrow} \sim O(e)$. There are two cases. If *c* and *e* do not actually occur, then the second counterfactual is automatically true because its antecedent and consequent are true: so *e* depends causally on *c* iff the first counterfactual holds. That is, iff *e* would have occurred if *c* had occurred. But if *c* and *e* are actual events, then it is the first counterfactual that is automatically true. Then *e* depends causally on *c* iff, if *c* had not been, *e* never had existed. I take Hume's second definition as my definition not of causation itself, but of causal dependence among actual events.

CAUSATION

Causal dependence among actual events implies causation. If *c* and *e* are two actual events such that *e* would not have occurred without *c*, then *c* is a cause of *e*. But I reject the converse. Causation must always be transitive; causal dependence may not be; so there can be causation without causal dependence. Let *c*, *d*, and *e* be three actual events such that *d* would not have occurred without *c* and *e* would not have occurred without *d*. Then *c* is a cause of *e* even if *e* would still have occurred (otherwise caused) without *c*.

We extend causal dependence to a transitive relation in the usual way. Let *c*, *d*, *e*, . . . be a finite sequence of actual particular events such that *d* depends causally on *c*, *e* on *d*, and so on throughout. Then this sequence is a *causal chain*. Finally, one event is a *cause* of another iff there exists a causal chain leading from the first to the second. This completes my counterfactual analysis of causation.

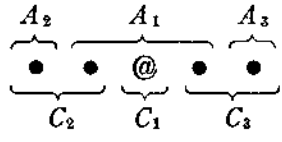
COUNTERFACTUAL VERSUS NOMIC DEPENDENCE

It is essential to distinguish counterfactual and causal dependence from what I shall call *nomic dependence*. The family $C_1, C_2, \ldots$ of propositions depends nomically on the family $A_1, A_2, \ldots$ iff there are a nonempty set $\mathfrak{L}$ of true law-propositions and a set $\mathfrak{T}$ of true propositions of particular fact such that $\mathfrak{L}$ and $\mathfrak{T}$ jointly imply (but $\mathfrak{T}$ alone does not imply) all the material conditionals $A_1 \supset C_1, A_2 \supset C_2, \ldots$ between the corresponding propositions in the two families. (Recall that these same material conditionals are implied by the counterfactuals that would comprise a counterfactual dependence.) We shall say also that the nomic dependence holds *in virtue of* the premise sets $\mathfrak{L}$ and $\mathfrak{T}$.

Nomic and counterfactual dependence are related as follows. Say that a proposition *B* is *counterfactually independent* of the family A_1,

A_2, . . . of alternatives iff B would hold no matter which of the A's were true—that is, iff the counterfactuals $A_1 \mathbin{\Box\!\!\rightarrow} B$, $A_2 \mathbin{\Box\!\!\rightarrow} B$, . . . all hold. If the C's depend nomically on the A's in virtue of the premise sets £ and ℱ, and if in addition (all members of) £ and ℱ are counterfactually independent of the A's, then it follows that the C's depend counterfactually on the A's. In that case, we may regard the nomic dependence in virtue of £ and ℱ as explaining the counterfactual dependence. Often, perhaps always, counterfactual dependences may be thus explained. But the requirement of counterfactual independence is indispensable. Unless £ and ℱ meet that requirement, nomic dependence in virtue of £ and ℱ does not imply counterfactual dependence, and, if there is counterfactual dependence anyway, does not explain it.

Nomic dependence is reversible, in the following sense. If the family $C_1, C_2, \ldots$ depends nomically on the family $A_1, A_2, \ldots$ in virtue of £ and ℱ, then also $A_1, A_2, \ldots$ depends nomically on the family $AC_1, AC_2, \ldots$, in virtue of £ and ℱ, where A is the disjunction $A_1 \vee A_2 \vee \ldots$. Is counterfactual dependence likewise reversible? That does not follow. For, even if £ and ℱ are independent of $A_1, A_2, \ldots$ and hence establish the counterfactual dependence of the C's on the A's, still they may fail to be independent of AC_1, $AC_2, \ldots$, and hence may fail to establish the reverse counterfactual dependence of the A's on the AC's. Irreversible counterfactual dependence is shown below: @ is our actual world, the dots are the other worlds, and distance on the page represents similarity "distance."



The counterfactuals $A_1 \mathbin{\Box\!\!\rightarrow} C_1$, $A_2 \mathbin{\Box\!\!\rightarrow} C_2$, and $A_3 \mathbin{\Box\!\!\rightarrow} C_3$ hold at the actual world; wherefore the C's depend on the A's. But we do not have the reverse dependence of the A's on the AC's, since instead of the needed $AC_2 \mathbin{\Box\!\!\rightarrow} A_2$ and $AC_3 \mathbin{\Box\!\!\rightarrow} A_3$ we have $AC_2 \mathbin{\Box\!\!\rightarrow} A_1$ and $AC_3 \mathbin{\Box\!\!\rightarrow} A_1$.

Just such irreversibility is commonplace. The barometer reading depends counterfactually on the pressure—that is as clear-cut as counterfactuals ever get—but does the pressure depend counterfactually on the reading? If the reading had been higher, would the pressure have been higher? Or would the barometer have been malfunctioning? The second sounds better: a higher reading would have been an incorrect reading. To be sure, there are actual laws and cir-

cumstances that imply and explain the actual accuracy of the barometer, but these are no more sacred than the actual laws and circumstances that imply and explain the actual pressure. Less sacred, in fact. When something must give way to permit a higher reading, we find it less of a departure from actuality to hold the pressure fixed and sacrifice the accuracy, rather than vice versa. It is not hard to see why. The barometer, being more localized and more delicate than the weather, is more vulnerable to slight departures from actuality.[10]

We can now explain why regularity analyses of causation (among events, under determinism) work as well as they do. Suppose that event c causes event e according to the sample regularity analysis that I gave at the beginning of this paper, in virtue of premise sets $\mathfrak{L}$ and $\mathfrak{F}$. It follows that $\mathfrak{L}$, $\mathfrak{F}$, and $\sim O(c)$ jointly do not imply $O(e)$. Strengthen this: suppose further that they do imply $\sim O(e)$. If so, the family $O(e)$, $\sim O(e)$, depends nomically on the family $O(c)$, $\sim O(c)$ in virtue of $\mathfrak{L}$ and $\mathfrak{F}$. Add one more supposition: that $\mathfrak{L}$ and $\mathfrak{F}$ are counterfactually independent of $O(c)$, $\sim O(c)$. Then it follows according to my counterfactual analysis that e depends counterfactually and causally on c, and hence that c causes e. If I am right, the regularity analysis gives conditions that are almost but not quite sufficient for explicable causal dependence. That is not quite the same thing as causation; but causation without causal dependence is scarce, and if there is inexplicable causal dependence we are (understandably!) unaware of it.[11]

EFFECTS AND EPIPHENOMENA

I return now to the problems I raised against regularity analyses, hoping to show that my counterfactual analysis can overcome them.

The *problem of effects*, as it confronts a counterfactual analysis, is as follows. Suppose that c causes a subsequent event e, and that e does not also cause c. (I do not rule out closed causal loops a priori, but this case is not to be one.) Suppose further that, given the laws and some of the actual circumstances, c could not have failed to

[10] Granted, there are contexts or changes of wording that would incline us the other way. For some reason, "If the reading had been higher, that would have been because the pressure was higher" invites my assent more than "If the reading had been higher, the pressure would have been higher." The counterfactuals from readings to pressures are much less clear-cut than those from pressures to readings. But it is enough that some legitimate resolutions of vagueness give an irreversible dependence of readings on pressures. Those are the resolutions we want at present, even if they are not favored in all contexts.

[11] I am not here proposing a repaired regularity analysis. The repaired analysis would gratuitously rule out inexplicable causal dependence, which seems bad. Nor would it be squarely in the tradition of regularity analyses any more. Too much else would have been added.

cause *e*. It seems to follow that if the effect *e* had not occurred, then its cause *c* would not have occurred. We have a spurious reverse causal dependence of *c* on *e*, contradicting our supposition that *e* did not cause *c*.

The *problem of epiphenomena*, for a counterfactual analysis, is similar. Suppose that *e* is an epiphenomenal effect of a genuine cause *c* of an effect *f*. That is, *c* causes first *e* and then *f*, but *e* does not cause *f*. Suppose further that, given the laws and some of the actual circumstances, *c* could not have failed to cause *e*; and that, given the laws and others of the circumstances, *f* could not have been caused otherwise than by *c*. It seems to follow that if the epiphenomenon *e* had not occurred, then its cause *c* would not have occurred and the further effect *f* of that same cause would not have occurred either. We have a spurious causal dependence of *f* on *e*, contradicting our supposition that *e* did not cause *f*.

One might be tempted to solve the problem of effects by brute force: insert into the analysis a stipulation that a cause must always precede its effect (and perhaps a parallel stipulation for causal dependence). I reject this solution. (1) It is worthless against the closely related problem of epiphenomena, since the epiphenomenon *e* does precede its spurious effect *f*. (2) It rejects a priori certain legitimate physical hypotheses that posit backward or simultaneous causation. (3) It trivializes any theory that seeks to define the forward direction of time as the predominant direction of causation.

The proper solution to both problems, I think, is flatly to deny the counterfactuals that cause the trouble. If *e* had been absent, it is not that *c* would have been absent (and with it *f*, in the second case). Rather, *c* would have occurred just as it did but would have failed to cause *e*. It is less of a departure from actuality to get rid of *e* by holding *c* fixed and giving up some or other of the laws and circumstances in virtue of which *c* could not have failed to cause *e*, rather than to hold those laws and circumstances fixed and get rid of *e* by going back and abolishing its cause *c*. (In the second case, it would of course be pointless not to hold *f* fixed along with *c*.) The causal dependence of *e* on *c* is the same sort of irreversible counterfactual dependence that we have considered already.

To get rid of an actual event *e* with the least over-all departure from actuality, it will normally be best not to diverge at all from the actual course of events until just before the time of *e*. The longer we wait, the more we prolong the spatiotemporal region of perfect match between our actual world and the selected alternative. Why diverge sooner rather than later? Not to avoid violations of laws of nature.

Under determinism *any* divergence, soon or late, requires some violation of the actual laws. If the laws were held sacred, there would be no way to get rid of e without changing all of the past; and nothing guarantees that the change could be kept negligible except in the recent past. That would mean that if the present were ever so slightly different, then all of the past would have been different—which is absurd. So the laws are not sacred. Violation of laws is a matter of degree. Until we get up to the time immediately before e is to occur, there is no general reason why a later divergence to avert e should need a more severe violation than an earlier one. Perhaps there are special reasons in special cases—but then these may be cases of backward causal dependence.

PREEMPTION

Suppose that c_1 occurs and causes e; and that c_2 also occurs and does not cause e, but would have caused e if c_1 had been absent. Thus c_2 is a potential alternate cause of e, but is preempted by the actual cause c_1. We may say that c_1 and c_2 overdetermine e, but they do so asymmetrically.[12] In virtue of what difference does c_1 but not c_2 cause e?

As far as causal dependence goes, there is no difference: e depends neither on c_1 nor on c_2. If either one had not occurred, the other would have sufficed to cause e. So the difference must be that, thanks to c_1, there is no causal chain from c_2 to e; whereas there is a causal chain of two or more steps from c_1 to e. Assume for simplicity that two steps are enough. Then e depends causally on some intermediate event d, and d in turn depends on c_1. Causal dependence is here intransitive: c_1 causes e via d even though e would still have occurred without c_1.

So far, so good. It remains only to deal with the objection that e does *not* depend causally on d, because if d had been absent then c_1 would have been absent and c_2, no longer preempted, would have caused e. We may reply by denying the claim that if d had been absent then c_1 would have been absent. That is the very same sort of spurious reverse dependence of cause on effect that we have just rejected in simpler cases. I rather claim that if d had been absent, c_1 would somehow have failed to cause d. But c_1 would still have been there to interfere with c_2, so e would not have occurred.

DAVID LEWIS

Princeton University

[12] I shall not discuss symmetrical cases of overdetermination, in which two overdetermining factors have equal claim to count as causes. For me these are useless as test cases because I lack firm naive opinions about them.